Amendments to the Specification:

Please amend the paragraph at page 1, lines 9-25 as follows:

The genomic DNAs of various living organisms are currently being sequenced and analyzed all over the world. The entire genomic sequences of more than 40 species of prokaryotes, a lower eukaryote, yeast, a multicellular eukaryote, *C. elegans*, and a higher plant, *Arabidopsis*, ~~*arabidopsis*~~, and such have already been determined. Analysis of the human genome, presumed to have three billion base pairs, was advanced under global cooperative organization, and a draft sequence was disclosed in 2001. In 2003 the complete structure had been elucidated and publically disclosed. A genome is a blueprint for highly complicated living organisms. The aim in determining a genomic sequence is to reveal the function and regulation of all genes, and to understand living organisms as a network of interactions between genes, proteins, cells or individuals. Understanding living organisms through the genomic information of various species is not only academically important, but also socially significant from the viewpoint of industrial application.

Please amend the paragraph at page 4, lines 18-28 as follows:

[1] SwissProt ~~(http://www.ebi.ac.uk/ebi_docsSwissProt_db/swisshome.html)~~,
[2] GenBank ~~(http://www.ncbi.nlm.nih.gov/web/GenBank)~~,
[3] UniGene (Human) ~~(http://www.ncbi.nlm.nih.gov/UniGene)~~,
[4] nr (a protein database, which has been constructed by combining data of coding sequences (CDS) in nucleotide sequences deposited in GenBank, and data of SwissProt, PDB ~~(http://www.rcsb.org/pdb/index.html)~~, PIR ~~(http://pir.georgetown.edu/pirwww/pirhome.shtml)~~, and PRF ~~(http://www.prf.or.jp/en/)~~; overlapping sequences have been removed.), and
[5] RefSeq ~~(http://www.ncbi.nlm.nih.gov/LocusLink/refseq.html)~~.

Please amend the paragraph at page 9, lines 15-34 as follows:

All of the full-length cDNAs of the present invention can be synthesized by a method such as PCR (Current protocols in Molecular Biology edit. Ausubel et al. (1987) Publish. John Wiley & Sons Section 6.1-6.4) using primer sets designed based on 5'-end and 3'-end sequences, or using primer sets of primers designed based on 5'-end sequences and a primer of oligo dT

sequence corresponding to poly A sequence. Table 1 contains the clone names of 2,495 full-length cDNA clones of the present invention, SEQ ID NOs of the full-length nucleotide sequences, CDS portions deduced from the full-length nucleotide sequences, and SEQ ID NOs of the translated amino acids. The positions of CDS are shown according to the rules set out in "DDBJ/EMBL/GenBank Feature Table Definition" ~~(http://www.ncbi.nlm.nih.gov/collab/FT/index.html)~~. The start position number corresponds to the first letter of "ATG", the nucleotide triplet encoding methionine; the termination position number corresponds to the third letter of the stop codon. These are indicated by flanking with the mark "..". However, in clones without a stop codon, the termination position is indicated by the mark ">", according to the above rules.

Please amend the paragraph bridging pages 85-86 as follows:

As used herein, "percent identity" of amino acid sequences or nucleic acids is determined using the BLAST algorithm of Karlin and Altschul (Proc. Natl. Acad. Sci. USA 90:5873-5877, 1993). Such an algorithm is incorporated into the BLASTN and BLASTX programs of Altschul et al. (J. Mol. Biol.215:403-410, 1990). BLAST nucleotide searches are performed with the BLASTN program, using for example, score = 100, wordlength = 12. BLAST protein searches are performed with the BLASTX program, using for example, score = 50, wordlength = 3. When utilizing the BLAST and Gapped BLAST programs, the default parameters of each program are used. ~~See http://www.ncbi.nlm.nih.gov.~~

Please amend the paragraph at page 93, lines 8-21 as follows:

First, a polynucleotide fragment of interest is inserted into the entry vector using the first recombination. Then, a second recombination is allowed to take place between the entry vector, where the polynucleotide fragment of interest has been inserted, and the destination vector. Thus, the expression vector can be prepared rapidly and efficiently. Using the above-mentioned typical restriction enzyme/ligase reaction method, expression vector construction and expression of a polypeptide of interest takes about seven to ten days. However, using the GATEWAY™ system, the polypeptide of interest can be expressed and prepared in only three to four days. Thus, the system ensures a high-throughput functional analysis for expressed polypeptides ~~(http://biotech.nikkeibp.co.jp/netlink/lto/gateway/)~~.

Please amend the paragraph at page 96, lines 5-16 as follows:

Alternatively, the function of a polypeptide encoded by a cDNA of the present invention can be predicted when a signal sequence, transmembrane domain, nuclear translocation signal, glycosylation signal, phosphorylation site, zinc-finger motif, SH3 domain, or such is found in the amino acid sequence. In particular, partial sequence structures such as motif and domain structures are commonly found in a number of proteins, and comprise a minimal functional protein structure. The Pfam database identifies a total of 4,832 types of motifs and domains, including both those whose functions have been clarified and those whose functions remain unclear ~~(http://www.sanger.ac.uk/Software/Pfam/index.shtml)~~ Version 7.7 (the latest version as of December 2002).

Please amend the paragraph at page 98, lines 5-22 as follows:

A more specific method for predicting function involves homology searches of databases such as GenBank, Swiss-Prot, UniGene, nr and RefSeq, using BLAST or FASTA. The functions of polypeptides encoded by the cDNAs of the present invention can be predicted based on hit genes and the function of polypeptides encoded by these genes. Polypeptide functions can be predicted from the amino acid sequences deduced from the structure of the full-length nucleotide sequences. In this way, signal sequences and transmembrane domains can be predicted from amino acid sequences using PSORT [K. Nakai & M. Kanehisa, Genomics, 14: 897-911 (1992)], SOSUI [T. Hirokawa et al., Bioinformatics, 14, 378-379 (1998)] (Mitsui Knowledge Industry Co., Ltd.), MEMSAT [D. T. Jones, W. R. Taylor & J. M. Thornton, Biochemistry, 33, 3038-3049 (1994)], and the like. Alternatively, motifs and domains can be predicted from amino acid sequences by carrying out searches using Pfam, PROSITE ~~(http://www.expasy.ch/prosite/)~~, or such. The above-described procedures facilitate more accurate prediction of polypeptide function.

Please amend the paragraph at page 128, lines 10-24 as follows:

Detailed descriptions concerning each domain or motif can be found in websites linked from the websites of Pfam, InterPro ~~(http://www.ebi.ac.uk/interpro/)~~, PROSITE ~~(http://www.expasy.ch/cgi-bin/prosite-list.pl)~~, or such. This information can be found based on domain/motif names, and accession numbers of hit data obtained through domain searches

of Pfam ~~(http://www.sanger.ac.uk/Software/Pfam/index.shtml)~~ (see Example 5) for amino acid sequences deduced from the 2,495 full-length clones of the present invention whose full-length nucleotide sequences have been determined. PROSITE in particular enables comparison of unique functional categories. The functions of polypeptides encoded by the 914 clones with hit data in Pfam were predicted and classified into the 13 functional categories described below. As a result, 661 clones were estimated to encode proteins belonging to these categories.

Please amend the paragraph at page 145, lines 15-29 as follows:

A clone predicted to belong to the category of disease-related protein means a clone having hit data in a homology search with some annotation, such as disease mutation and syndrome, suggesting that the clone encodes a disease-related protein; or a clone whose full-length nucleotide sequence has hit data in Swiss-Prot, GenBank, UniGene, nr or RefSeq, where that hit data corresponds to genes or polypeptides which have been deposited in the Online Mendelian Inheritance in Man (OMIM) ~~(http://www.ncbi.nlm.nih.gov/Omim/)~~, the human gene and disease database described later; or a clone in which the results of a domain/motif search with Pfam suggest the presence of domains and motifs, that suggest proteins with disease-specific expression or proteins involved in increasing or decreasing expression (depending on the disease), for example, Wilm's tumor protein or von Hippel-Lindau disease tumor suppressor protein.

Please amend the paragraph at page 151, lines 11-27 as follows:

There are several methods for analyzing the expression level of genes involved in disease. Differences in gene expression levels between diseased and normal tissues can be studied by analytical methods using, for example, Northern blotting, RT-PCR, DNA microarrays, etc. (Experimental Medicine, Vol.17, No. 8, 980-1056 (1999); Cell Engineering (additional volume) DNA Microarray and Advanced PCR Methods, Muramatsu & Nawa (eds.), Shujunsya (2000)). In addition to these analysis methods, computer analysis can be used to compare the nucleotide sequences of expressed genes, and hence to analyze expression frequency. For example, in the "BODYMAP" database, gene clones are randomly extracted from cDNA libraries of various tissues and/or cells, clones homologous to each other are

assigned to a single cluster based on 3'-end nucleotide sequence homology information, genes are then classified into clusters, and the number of clones in each cluster is compared to gain information on expression frequency (~~http://bodymap.ims.u-tokyo.ac.jp/~~).

Please amend the paragraph bridging pages 216-217 as follows:

With respect to the amino acid sequences deduced from the full-length nucleotide sequences, the prediction was made for the presence of signal sequence at the amino terminus, the presence of transmembrane domain, and the presence of functional protein domains (motifs). The signal sequence at the amino terminus was searched for by PSORT [K. Nakai & M. Kanehisa, Genomics, 14: 897-911 (1992)]; the transmembrane domain, by SOSUI [T. Hirokawa et al., Bioinformatics, 14: 378-379 (1998)] (Mitsui Knowledge Industry); the function domain, by Pfam (Version 5.5) (~~http://www.sanger.ac.uk/Software/Pfam/index.shtml~~). The amino acid sequence in which the signal sequence at the amino terminus or transmembrane domain had been predicted to be present by PSORT or SOSUI were assumed to be a secretory or membrane protein. Further, when the amino acid sequence hit a certain functional domain by the Pfam functional domain search, the protein function can be predicted based on the hit data, for example, by referring to the function categories on the PROSITE (~~http://www.expasy.ch/cgi-bin/prosite-list.pl~~). In addition, the functional domain search can also be carried out on the PROSITE.

Please amend the paragraph at page 270, lines 15-24 as follows:

A clone predicted to belong to the category of disease-related protein means a clone having hit data with some annotation, such as disease mutation and syndrome, suggesting that the clone encodes a disease-related protein, or means a clone whose full-length nucleotide sequence has hit data for Swiss-Prot, GenBank, or UniGene, where the hit data corresponds to genes or proteins which have been deposited in the Online Mendelian Inheritance in Man (OMIM) (~~http://www.ncbi.nlm.nih.gov/Omim/~~), which is the human gene and disease database.

Please amend the paragraph at page 309, lines 9-17 as follows:

Pfam was used to undertake a domain search for the amino acid sequences deduced from the full-length nucleotide sequences (see Example 5). Based on these results, the proteins encoded by clones 664 and 250 were categorized and their functions predicted. This was performed by referring to domain and motif names, accession numbers for hit data, and detailed descriptions in Pfam ~~(http://www.sanger.ac.uk/Software/Pfam/index.shtml)~~ as well as functional categorizations in PROSITE ~~(http://www.expasy.ch/cgi-bin/prosite-list.pl)~~.

Please amend the paragraph at page 439, line 33 as follows:

~~BRACE2035371~~ BRACE2035381 10.177 10.254

Please amend the paragraph at page 519, lines 7-34 as follows:

Various methods for analyzing gene expression frequency have been developed. For example, wet-type experiment-based methods include Northern blotting and RT-PCR, and of the use of gene chips and microarrays, where target samples synthesized from tissue or cell-derived RNA are hybridized to polynucleotides that comprise partial gene sequences synthesized on a base, or cDNA clones attached as plasmids directly to a base, and then signals are detected (Experiment Medicine, Vol. 17, No. 8, 980-1056 (1999), Eds., Muramatsu and Nawa, Cell Technology, Suppl. "DNA Microarray and New PCR Methods" (Shujunsha, 2000)). A method called "ATAC-PCR" is also available (Kato. K (1997) Nucleic Acids Res. 25, 4694-6), which comprises the steps of cleaving cDNA synthesized from tissue or cell-derived RNA, attaching adapters of different length depending on the type of tissue or cell, carrying out competitive PCR using a primer which contains a fluorescent dye and a sequence complementary to the adapter, and a primer specific to the gene, and then analyzing the expression level of the gene. In addition, an in-silico analysis-based method using sequence data is available. A database called BODYMAP ~~(http://bodymap.ims.u-tokyo.ac.jp/)~~ has been constructed by randomly extracting gene clones from cDNA libraries of various tissues and cells, combining clones homologous to one another as a cluster, classifying the genes in each cluster unit based on homology information on the nucleotide sequences of cDNA 3' ends, and then obtaining information on gene expression frequency by comparing the number of clones in respective clusters.